Robust Vision-based Underwater Target Identification & Homing Using Self-Similar Landmarks

Amaury Nègre, Cédric Pradalier, Matthew Dunbabin

Robust Vision-based Underwater Target Identification & Homing Using Self-Similar Landmarks

Amaury Negre[1], Cedric Pradalier[2] and Matthew Dunbabin[2]

[1] INRIA Rhônes-Alpes, France - firstname.lastname@inrialpes.fr
[2] Autonomous Systems Laboratory, CSIRO ICT Centre, Australia
firstname.lastname@csiro.au

Abstract. Next generation Autonomous Underwater Vehicles (AUVs) will be required to robustly identify underwater targets for tasks such as inspection, localisation and docking. Given their often unstructured operating environments, vision offers enormous potential in underwater navigation over more traditional methods, however, reliable target segmentation often plagues these systems. This paper addresses robust vision-based target recognition by presenting a novel scale and rotationally invariant target design and recognition routine based on Self-Similar Landmarks (SSL) that enables robust target pose estimation with respect to a single camera. These algorithms are applied to an AUV with controllers developed for vision-based docking with the target. Experimental results show that system performs exceptionally on limited processing power and demonstrates how the combined vision and controller systems enables robust target identification and docking in a variety of operating conditions.

1 Introduction

Target identification and homing are of particular interest in our research as we desire an Autonomous Underwater Vehicle (AUV) fitted with a cam-

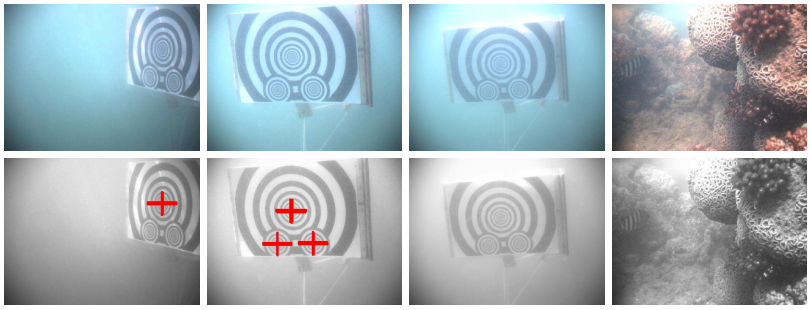

Fig. 1. SSL detection in marine environment. Upper row: original images, lower row: detection. From left to right: robustness to perspective and lighting discontinuity, minimum required contrast, insufficient contrast, absence of false positives in a natural environment.

era and vision processing capabilities to accurately locate and home onto points of interest in relatively cluttered environments such as coral reefs, or in close proximity to subsea oil and gas structures. Due to issues such as multi-pathing and variable lighting, the performance of typical identification and homing methods such as acoustics or feature and color-based vision systems can degrade significantly. The ability to robustly identify objects in these environments can allow tasks such as reliable localisation, inspection, object collection and docking to complex structures.

Underwater targets for docking/homing can be either active or passive and are typically identified using either acoustics or vision systems. Active (energy emitting) targets are most common in literature due to their generally larger detection range, whereas passive targets can be further classified as artificial or natural and are typically detected using vision systems at closer distances.

Acoustics and vision are the most common ways of identifying and homing onto targets. Acoustics are advantageous in that they can operate over long ranges and in a variety of water conditions (visibility and lighting). However, for docking, acoustics degrade at very close ranges (< 5m) and in cluttered environments such reefs or close to the seafloor and subsea structures. Vision-based target identification has complimentary properties. Vision is suitable for close range tracking and can accommodate changing and multiple targets, however, its performance degrades in turbid water and poor lighting conditions.

There are many examples of target identification and homing in the literature. The simplest approach is to home to an acoustic target using range and heading information from the target [8]. Although a reliable method at larger distances, at close ranges the performance becomes impractical for docking due to the high update rates required. Alternatively, Feezor [5] successfully demonstrated homing and docking using electromagnetic guidance in which an AUV was fitted with coils to sense field strength and orientation that enabled docking from a distance of 30m.

Improvements in the docking performance of AUVs have been considered by combining acoustics with vision [4]. Here, the acoustics/sonar provides the longer range homing direction with the vision providing guidance in the final stages of docking. A survey of vision-based target identification and tracking literature is provided by Dalgleish [2] describing the optic flow and feature based techniques commonly used for underwater tracking applications.

Many vision-based target identification schemes require active targets. Lee [7] uses a monocular vision system to identify a large circular target with an LED ring and 5 large lights, whereas Wang [9] describes a stereo vision system to locate and retrieve an underwater object which has a light emitting beacon on its surface.

Considering all these methods of target identification and homing, it was decided that our system would require passive targets due to energy requirements on behalf of the AUV and target itself for long duration operations.

Additionally, when repeat visitations of a target over extended periods of time are required, natural landmarks can change significantly, requiring artificial landmarks. However, detection is required at relatively large distances and different orientations making color-based methods unreliable. Therefore, the target and identification routine must be robust, scale and rotationally invariant, as well as capable of running in real-time on an AUV's limited processing power. This paper considers the use of Self-Similar Landmarks (SSL) [1] as a robust, color, scale and rotational invariant means of target identification from which the information can be used to guide and AUV for homing and docking operations.

2 Self-Similar Landmarks

The notion of self-similar landmarks was first used in a robotic context by Scharstein and Briggs [1]. Their objective was to develop planar targets that would be detected easily with a standard perspective camera on a mobile indoor robot. To this end, the targets were designed to be invariant to change of scale. This is where the self-similarity is essential.

A p-similar function for $0 < p < 1$ is a function $f : \mathbb{R}^+ \to \mathbb{R}$ such that $\forall x > 0, f(x) = f(p \cdot x)$. It is essential to note that the p-similarity is invariant to change of scale ($\forall x > 0, \forall k > 0, f(k \cdot x) = f(p \cdot k \cdot x)$). In the context of a computer vision application, this property is interesting since, if a method to detect self-similarity is available, it will be possible to do so from whatever distance since, with a pin-hole camera, a change of range results in a change of scale in the observed image.

When designing a self-similarity detector, the fact that all constant functions are p-similar for all p is problematic since any uniform region in an image will give a string response on the self-similarity detector. To solve this, [1] introduces the notion of anti-similarity: a function f is p-antisimilar if $\forall x > 0, |f(x) - f(p \cdot x)| = 1$.

The p-similarity and $\sqrt{p}$-antisimilarity are then used by Scharstein and Briggs to define the "Self-similar square wave" function which is both p-similar and $\sqrt{p}$-antisimilar:

$$s_p(x) = \lfloor 2(\log_p x - \lfloor \log_p x \rfloor) \rfloor \tag{1}$$

This function maps $\mathbb{R}^+$ to $\{0, 1\}$. It is used to print black and white landmarks, as shown in fig. 2(a).

To recognise the self-similar landmarks, [1] defines a matching function that will respond strongly to a sequence of pixels looking like a p-similar and $\sqrt{p}$-antisimilar function. For a pixel located at (x, y) in image I, this is achieved by integrating the p-similarity and $\sqrt{p}$-antisimilarity conditions on

a window of length w starting at (x, y):

$$m^w_{+x}(x,y) = \frac{1}{w}\int_0^w |I(x+\xi,y) - I(x+\sqrt{p}\cdot\xi,y)|\,d\xi - \frac{1}{w}\int_0^w |I(x+\xi,y) - I(x+p\cdot\xi,y)|\,d\xi \quad (2)$$

In this equation, the window is made of the w pixels after (x, y), on an image line in the x-direction. We identify this function by the $+x$ subscript. The function can be obviously modified to use the w pixels before, or vertically above, or below (x, y). We name these functions m^w_{-x}, m^w_{-y} and m^w_{+y} respectively. In practical implementations, w takes values in the $[20, 100]$ pixel range. As in [1], we use 40 as a good balance between robustness and computation cost.

2.1 A Modified SSL for Pose & Range Estimation

The two-dimensional landmark used by Briggs is self-similar on one dimension (horizontal for example) and constant on the other one. This landmark is not optimal for many applications, particularly underwater:

- This is not a point landmark and consequently it is hard to localise precisely in the image. In [1], the border of the landmark is detected, and a binary coding is used to identify the landmark. Our practical experience in outdoor experiments found this to be the weakest part of the algorithms.
- This brings us to our second point, this landmark is not robust to rotations. Since only the m^w_{+x} function is used, the landmark can only be detected when the apparent rotation is below 45 degrees (from our experiments).
- Also, this landmark is not robust to motion blur: a blur in the direction of the p-similar function can make the landmark detection fail.

In order to solve those problems, we designed a circular landmark where the intensity I is self-similar and anti-similar in all directions:

$$I(\rho,\theta) = s_p(\rho) \quad (3)$$

where ρ, θ are the polar coordinate and s_p the self-similar square wave function defined in Equation 1.

This landmark is interesting because it is rotationally invariant thanks to the circular geometry and the matching function exhibits a single maximum point in the center. Moreover, to improve robustness to noise and motion blur, we can apply the matching function in several directions (e.g. top, bottom, left and right).

(a) Original SSL target (b) Circular 3-pattern SSL target. (c) SSL detector output

Fig. 2. The original(a) and circular SSL(b) target patterns used in this investigation and the output(c) of our circular SSL detector

In our application we also need to estimate the range and the pose of the target. Range could be estimated using a single circular landmark and stereo camera system. However, detection of distant objects requires large camera baselines and this also doubles the image processing requirements. In this research, a monocular vision was chosen to reduce computation burden and allow portability to other platforms. As we can detect only the center of the pattern we need at least 3 patterns to estimate the pose[1]. Our final target (shown in Figure 2(b)) consists of 3 circular self-similar patterns on an equilateral triangle: one large pattern for detection at greater distances and two small patterns for the pose estimation at short range.

3 SSL Target Identification & Tracking

3.1 Target Detection

The first step in the detection algorithm consists of applying the matching function on every pixel of the image in four perpendicular directions:

$$M^w(x,y) = m^w_{+x}(x,y) + m^w_{-x}(x,y) + m^w_{+y}(x,y) + m^w_{-y}(x,y) \qquad (4)$$

Next, the location of the local maxima of the matching function is determined with a threshold function applied to remove small outlier peaks.

The results and a typical response of the matching function are shown in Fig. 2(c). In our testing, this target appeared robust to occlusion (as long as the center point is visible), deformation resulting from bending or shaking the paper landmark, camera model (the same detection software worked unmodified on various focal length pin-hole cameras in air or water, and also on fish-eye and catadioptric cameras) and lighting change. For underwater detection, the only possible problematic visual effects are water turbidity and the reflection of the landmark by the water-air interface.

[1] It is well known that at most 8 poses will be consistent with the observation[6].

3.2 Pose & Range Estimation

Once the position of the landmarks in the image has been found, we then need to evaluate the pose of the target with respect to the camera. Three cases are possible according to the number of visible landmarks.

Case 1: only one visible landmark When the target is far from the camera, we typically detect only the large landmark. Therefore, when using a monocular camera, only the bearing can be determined.

Case 2: two visible landmarks In the case where only two landmarks are visible, there is no reliable way to detect which ones are seen. Consequently, we treat this case as if only the larger landmark was visible.

Case 3: three visible landmarks When the target is close enough, all landmarks are visible and it is possible (with some ambiguities) to compute the pose from three points.

3.3 Target Tracking

The underwater environment and the motion of the submarine make the target detection subject to noise and misdetection. In order to be less sensitive to these problems, we developed a target tracking system based on a particle filter (see [3]).

Each particle represents a pose of the target in the robot reference, coded by a 7-dimensions vector: the first four coordinates represent the rotation as a quaternion and the last three coordinates represent the 3D position of the centroid of the target.

The particle filter is initialized when the system detects at least one landmark for the first time (or after a long period). We then initialise all particles around the position estimated as explained in Section 3.2. For the prediction phase of the particle filter, we use the odometric sensors (accelerometers, compass, and depthmeter) to update the particles with the estimated motion blurred with a gaussian noise. In the updating phase, the weight of the particles is estimated by projecting the virtual landmarks represented by the particle in the image and evaluating the distance to the observed landmark.

4 Practical Limitations of SSL in Field Robotics

Using SSL is computationally costly. The detection requires $O(n^2w)$ non-sequential image accesses, where n is the image size and w the size of the integration window. The non-sequential pixel indexing aspect is critical for an optimized implementation but it prevents an efficient implementation using specialised processor instruction such as *MMX*, *SSE* or *SSE2*[2] and also prevents efficient caching of image data.

[2] Our comparison of implementation using floating point operations, integer operations, MMX, SSE and SSE2 showed compiler optimised floating point implementation to be the most effective on our platform.

4.1 Improving Performance

As mentioned previously, the processor requirement is important to compute the self-similar matching function in real-time. Table 1 lists the measured performance of the proposed target identification technique on different computing platforms and image sizes.

To improve performance, a simple way is to only compute this function in a reduced Region Of Interest (ROI) in the image. Using the particle filter, we can predict the position of the 3 landmarks in the image and reduce the search region around these 3 positions. Typically, in our application the ROI is a 100x100 pixel rectangle around each predicted point, dividing the computation time by 3 for an image 640x480 (see Table 1). Nevertheless, to avoid the risk of losing the target, we maintain a search in the whole image at least every 10 frames.

Processor	image size	FPS for whole image	FPS with tracking (ROI 100x100 pixels)
Pentium IV 3GHz	640x480	2.8	8.7
	320x240	11.1	18.7
Pentium M 1.4GHz	640x480	2.1	6.8
	320x240	8.9	15.4

Table 1. Computation time for different processor and different image sizes.

5 Experimental Results

The vision-based target identification algorithm presented above was evaluated on an AUV in semi-controlled (test tank) conditions. The purpose of these experiments was to evaluate the robustness and performance of the circular SSL tracking system, as well as the AUV's target identification and homing/docking capabilities under different operating scenarios.

The Starbug AUV was the platform used for these experiments. Starbug, shown in Fig. 3, uses vision as its primary sensory mode for navigation. The vision system consists of two stereo camera pairs, one looking forward along the x-axis for obstacle avoidance, and the other looking downwards along the z-axis for visual odometry and terrain following.

As the SSL algorithm requires only monocular vision, only one of the forward (x-axis) cameras was used in these experiments. Additionally, the roll, pitch and yaw measurements from the on-board IMU were made available to the target tracking algorithm. All visual and inertial measurements are shared amongst all the AUV's sub-systems using the DDX middleware.

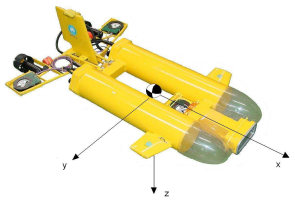

Fig. 3. The Starbug AUV used for docking experiments showing local coordinate system.

The target consisted of the circular SSL geometry (Fig. 2(b)) printed on A3 paper which was laminated and glued to a rigid backing. Anchors and floats attached to the target enabled the distance from the sea-floor to be set, however, it could move with water currents.

5.1 Target homing and stand-off experiments

The experiments consisted of developing a target searching routine for the AUV and when detected, use the vision-based SSL tracking algorithm to guide (home) the AUV towards the target and maintain a stand-off distance from it.

The AUV target identification and docking behaviour consisted of three states; State(1) is the target search mode whereby the AUV performs a constant rate spin about its z-axis at varying depths. The AUV enters State(2) when the target has been identified and remains until the AUV's x-axis is pointing at the identified target's centre. Finally, State(3) is the homing/docking mode where the AUV moves towards the target and maintains a prespecified distance from the target.

Figure 4 shows results of a typical autonomous target identification and homing experiment in which it is desired to maintain a distance of 2m from the target. Here the target was placed at one end of the test tank and the AUV at the other. Figure 4 shows the target tracking state as well as demanded (dashed) and actual (solid) AUV yaw angle and depth. Additionally, the number of SSL patterns found (nbPoints) in the image is shown along with the estimated range from the target identification algorithm (see Section 3).

From Figure 4, it can be seen that the target is first identified at t=75s and as the AUV moves towards the target, the 2m standoff distance is maintained. At t=140s, the target was removed from the water and the AUV re-entered the search mode until t=225s where the target was replaced and the AUV reacquired and tracked it. This experiment was repeated many times with the system consistently able to identify and home onto the target.

To demonstrate the robustness of the target tracking, during maintaining the 2m standoff distance shown in Fig. 4, the target was moved left then right before being returned to its original position. Figure 5 shows the AUV yaw angle as well as the estimated target position relative to the AUV coordinate frame. Here the lateral (y_{rel}) position of the target moves as the target moves left and right and the yaw angle varies to maintain the target directly in front of the AUV.

Finally, the ability of the system to dock with the target is demonstrated in Fig. 6 by setting the stand-off distance to zero. The figure shows the actual and demanded AUV yaw and depth, as well as the estimated range to the target and number of SSL circles tracked. Additionally, the ADC value of

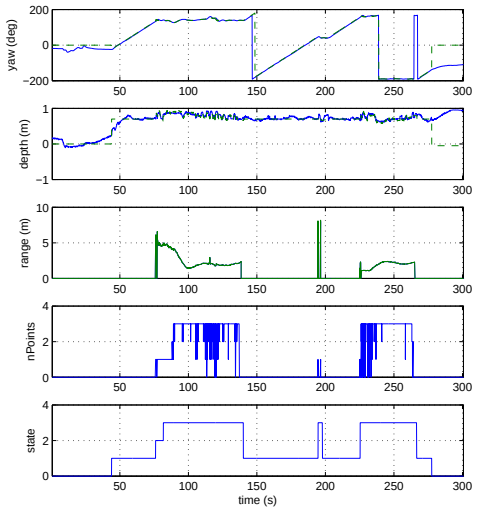

Fig. 4. AUV pose and SSL detection properties during a typical target identification and homing experiment. Here the AUV was required to identify the target and maintain a stand-off distance (range) of 2m.

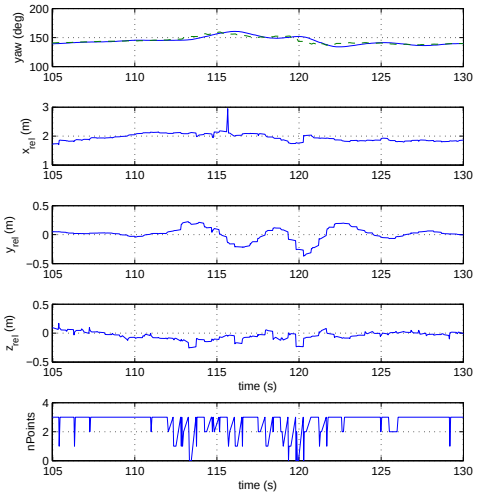

Fig. 5. Target tracking performance showing target position relative to AUV coordinate frame moving the target left and right and orienting it relative to the AUV.

the frontal collision avoidance sensor is shown. A collision is detected when the ADC value exceeds 300. The AUV detects a collision with the target at t=511s. Again this experiment was repeated several times with consistent performance demonstrated.

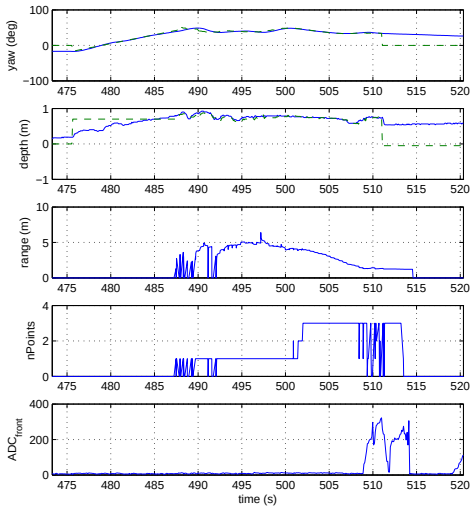

Fig. 6. AUV pose and SSL detection properties during a typical docking experiment where the AUV identifies and moves towards the target until the frontal collision sensor was triggered.

6 Conclusions

Underwater vision-based tasks are typically complex to implement due to poor lighting conditions, refractions and moving objects such as fish. In this paper we developed a rotationally invariant circular self-similar landmark and demonstrated its use for target identification and in enabling vision-based docking using the Starbug AUV. The method provides an exceptionally robust landmark with very little sensitivity to camera model, distortion and observation range. The resulting docking task was proven effective through extensive pool trials. In May 2007, an experiment in uncontrolled reef environment has been conducted to demonstrate the applicability of our approach in the field. The SSL detection proved effective (see fig. 1), but our control was not reactive enough to perform the docking in presence of current. This problem will be approached in future research.

References

1. A. Briggs, D. Scharstein, D. Braziunas, C. Dima, and P. Wall. Mobile robot navigation using self-similar landmarks. In *Proc. International Conference on Robotics and Automation ICRA '00*, pages 1428–1434, 2000.
2. F. Dalgleish, S. Tetlow, and R. Allwood. Vision-based navigation of unmanned underwater vehicles: A survey. part 2: Vision-based station-keeping and positioning. *Journal of Marine Design and Operations*, pages 13–19, 2005.
3. A. Doucet, N. De Freitas, and N. Gordon, editors. *Sequential Monte Carlo methods in practice*. 2001.
4. J. Evans, P. Redmond, C. Plakas, K. Hamilton, and D. Lane. Autonomous docking for Intervention-AUVs using sonar and vision-based real-time 3D pose estimation. In *Proc. OCEANS 2003*, volume 4, pages 2201–2210, September 2003.
5. M. Feezor, Y. Sorrell, P. Blankinship, and J. Bellingham. Autonomous underwater vehicle homing/docking via electromagnetic guidance. *Journal of Oceanic Engineering*, 26:515–521, 2001.
6. R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. 2000.
7. P.-M. Lee, B.-H. Jeon, and S.-M. Kim. Visual servoing for underwater docking of an autonomous underwater vehicle with one camera. In *Proc. OCEANS 2003*, volume 2, pages 677–682, September 2003.
8. R. Stokey, M. Purcell, N.Forrester, T. Austin, R. G. aand B. Allen, and C. van Alt. A docking system for REMUS, an autonomous underwater vehicle. In *Proc. Oceans '97*, pages 1132–1136, October 1997.
9. H. Wang, S. Rock, and M. Lee. Experiments in automatic retrieval of underwater objects with an AUV. In *Proc. OCEANS '95 MTS/IEEE*, pages 1–8, San Diego, October 1995.